Quantifying Visual Similarity in Clinical Iconic Graphics

Justin B. Starren, M.D. Ph.D.; Philip R.O. Payne, M.A.
Department of Biomedical Informatics, Columbia University, New York, New York

Abstract: Numerous studies have demonstrated the benefits of utilizing iconic presentation models in the context of complex medical information. However, very little literature exists that addresses the design of the graphical primitives that constitute such mediums. Utilizing a method named "Presentation Discovery", the authors of this study examine a manner in which objective techniques may be utilized to prototype such graphical primitives in order to increase the realized expressiveness of ensuing iconic presentation models.

Numerous studies have illustrated the benefits of iconic presentations for medical concepts. Metaphor icons are designed to be easier to interpret by somehow "looking like" the underlying concept being presented[1]. The interpretation of such icons is based on the assumption that different users perceive the shapes similarly. At the most basic level, different users must agree on which shapes are similar and which are different. If this similarity can be measured, it may be usable as the basis to design better icons.

The authors of this study examined such a measurement methodology, named "Presentation Discovery." In this particular instance, Six radiologists were asked to draw graphics representing 50 unique findings found in mammography reports, extrapolated using the MedLEE natural language processing system[2], resulting in 237 total graphics. A separate group of subjects, without domain knowledge, were then asked to sort these graphics based on the similarity of their shape, size, and style. The results of the observers sorting behavior was recorded in a 237 by 236 element matrix. Each possible pairing of graphics was assigned a score from 0-6 based on the number of times observers agreed on the grouping of two distinct graphics.

Using this data, aggregate user agreement for each potential pairing of a single graphics with the remaining 236 graphics was calculated, and found to be 68% with an SD of 10.85%. Semantic agreement was analyzed by calculating the mean of the aggregate user agreement for each of the 16 unique combinations that could be generated from a set of six graphics that were initially derived from a common textual finding, and found to be 74% with an SD of 16.5%.

Due to the methods utilized in this evaluation, each observer created a different number of groups, each with a unique number of members, indicating that conventional statistical measures of inter-observer agreement were not applicable. In order to assess the significance of the scores obtained, an algorithm that incorporated variables representing both the total number of graphical elements being sorted, as well as the average size of the groups generated by the observers, was designed. The resulting measurement, referred to as a threshold of significance, was found to be 3.8%. This algorithm was verified by performing a computer-simulation which randomly assigned agreement scores to an identical matrix, which when analyzed over 100 iterations, indicated an aggregate agreement score of 3.98% with a SD of 0.24%.

Further analysis was then performed to develop a "consensus clustering" of the groups of graphics created by the multiple observers. An unweighted pair-group centroid hierarchical clustering algorithm was applied to the initial matrix of scores[3]. The results of this analysis are as follows:

Cluster Type	Number of Clusters	Average Size	Euclidian Distance
Individual Graphics	89	4.35	4.22
Semantic Groups	17	3.88	7.25

Given these results, in can be concluded that observers reliably agreed upon the composition of groups of primitives, both at the individual graphic, and semantic levels. Furthermore, clustering analysis demonstrated the existence of multiple, "tightly" clustered groups of graphics. Given these observations, it can be postulated that different individuals will in fact categorize simple iconic graphics into similar clusters, and that these clusters may provide the elemental units for the construction of more intuitive icons.

References

1. Cole WG. Metaphor Graphics & Visual Analogy for Medical Data. Annual Symposium on Computer Applications in Medical Care; 1987; Washington D.C., United States of America.
2. Friedman C. Towards a Comprehensive Medical Language Processing System: Methods & Issues. Proceedings of the AMIA Annual Fall Symposium; 1997; Washington D.C., United States of America.